

Event Detection in Evolving Networks

Sarvenaz Choobdar, Pedro Ribeiro, and Fernando Silva

CRACS and INESC-TEC

Faculdade de Ciencias, Universidade do Porto, Portugal

Email:{sarvenaz,pribeiro,fds}@dcc.fc.up.pt

Abstract—This paper describes a methodology for finding and describing significant events in time evolving complex networks. We first group the nodes of the network in clusters, according to their similarity in terms of a set of local properties such as degree and clustering coefficient. We then monitor the behavior of these groups over time, looking for significant changes on the size of the groups. These events are notable since they show that the position of a number of nodes in the network has changed. We describe this evolution by extracting the correspondent transition patterns. We examined our methodology on three different real network datasets. Our experiments show that the discovered rules are significant and can describe the occurring events.

Index Terms—Network Characterization, Node labeling, Clustering, Cluster Evolution

I. INTRODUCTION

Advances in information technology led the world activity to become very much centered on information data. The explosive growth in data that we are witnessing naturally opens an enormous opportunity for researchers to develop new methodologies to dynamically extract useful information and knowledge from the data. Real life data inherently contains structural information on objects and their relationships. This structure can be modeled with networks, or graphs, that are abstract representations of a set of nodes and the connections between them.

Most real world networks are complex, in the sense that they present non trivial topological features. Research on complex network data analysis has been very prolific and a large variety of characterization methodologies emerged, such as graph clustering [1], [2], node classification [3], network motifs discovery [4] or frequent subgraph mining [5], [6]. These approaches treat the network as a static object. However, many networks are intrinsically dynamic and change over time. Only recently has the research community started to analyze the temporal evolution of networks [7], [8]. Most of these studies have characterized network structure by directly examining the topology of the network. Nevertheless, more indirect methods that use network measurements such as degree centrality or clustering coefficient can be a rich source of information [9]. Our aim is precisely to use the evolution of these kind of metrics to study the dynamics of networks that change over time.

We propose a novel two-phase general methodology designed to characterize time evolving networks. The first step of this methodology mainly follows our previous work on node label acquisition where we group and classify nodes based on

their role in the network [10]. In the second step we propose a different method to study the evolution of the network by a supervised approach. A set of events happening in the network are defined and discovered in the network. We then find the predefined events happening in the network and the rules that describe them.

The first phase involves looking at the network from a static point of view and creating a node classification. We can either use a predefined label for each node, or an automated classification of nodes based on their local properties, such as degree or betweenness centrality, which have been shown to be very fruitful in node characterization [9]. It is however not an easy task to choose a set of these kind of metrics that best generally describes and distinguishes nodes. Costa et al. [11] presented a node label acquisition methodology based on these type of metrics, but they only identify outliers, that is, singular nodes that are most different from the others. Instead, we apply these measurements to all nodes in the network and then employ clustering techniques to group nodes and assign them labels. With this we create distinct groups of nodes that are representative of the whole network.

The second phase involves finding significant events in the network. We define five event categories: growth, shrink, emerge, dissolve and constant. Each event is described by the time interval of its occurrence and the pattern that shows the origin or destination of nodes in every event.

We apply our methodology on three real complex networks to demonstrate the validity and usefulness of our approach. We show that our method can discover interesting insights in dynamic networks. In addition, the validation of the resulted pattern is examined by calculating the respective z-score.

In the remainder of the paper, we start by reviewing the state of the art work regarding network evolution. After, we describe the proposed methodology in section III, with all intermediate steps and techniques used. Then, we evaluate our approach in section IV, by applying it to network datasets. Finally, we draw some conclusions.

II. RELATED WORK

A. Graph clustering

As defined by Satu E. Schaeffer [2], “graph clustering is the task of grouping the nodes of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters”. Most of the existing graph clustering methods use only the topological structure of

a graph to partition it into the cohesive clusters. Clustering based on normalized cut [12], modularity [13] or structural density [14] are some of the common approaches.

On the other hand, there are some more recent studies that partition the graph according to its local properties in a way that nodes with the same feature vector are grouped into one partition [15], [16].

The selection of a distance measure differentiates these two approaches from each other. In the first group similarity of nodes is measured based on connectivity (e.g., the number of possible paths between two nodes) and structural similarity (e.g., the number of common neighbors of two nodes); while in the second group the similarity of two nodes is measured by feature similarity (e.g., Euclidian distance between two feature vectors).

B. Label acquisition

Label acquisition, as most commonly defined in the literature, involves determining the label for a node in a network that is partially labeled. Normally, it is assumed that at least some of the nodes have a predefined label and only the labels for remaining nodes are predicted using relational classifiers [17].

With networked data, the label of a node may influence the label of a related node. Furthermore, nodes not directly connected may be related through chains of links. This complex dependencies thus suggest that it may be beneficial to predict the label of all nodes simultaneously. Regarding the values of an attribute or attributes for multiple connected nodes for which some attribute values are unknown, a simultaneous statistical assessment is required and this can be done by using collective inferencing [18]. Networked data allow collective inference, meaning that various interrelated values can be derived simultaneously [3]. Gallagher and Eliassi-Rad [19] used another source of information in networks that is independent of the available node labels and improved the accuracy of node's label by adding label independent features which include nodes local measurements like degree and betweenness.

All the mentioned studies aim to find the label of a node in a partial labeled network and rely mainly on the available information about label of some of the nodes, they predict labels of nodes instead of assigning the labels. Our work in label acquisition follows the work by Costa et al. [11], but differs in that we consider all the nodes in the network instead of just the singular node-motifs.

C. Node evolution

Different approaches for explaining network evolution have been reported in the literature. Some have focused on the global evolution of networks by an exploratory point of view. [7] discovered the shrinking diameter phenomena on time-evolving networks. [20] studied the evolution of communities in social networks. Still from an exploratory perspective, [21] studied the evolution of networks but at a more local level.

Using a methodology based on the maximum-likelihood principle, [8] investigates a wide variety of the network formation strategies, and shows that edge locality plays a critical role in evolution of networks.

[22] described node's centrality changes over time and showed that hubs do not remain a hub for the all time. They use nodes degree over time to compute correlations between pairs of daily networks.

Other recent papers present algorithmic tools for the analysis of evolving networks. [23] focuses on assessing the community affiliation of users and how this changes over time. [24] applies the minimum description length (MDL) principle to the discovery of communities in dynamic networks, developing a parameter-free framework. This is the main difference to previous work such as [25], [26]. However, as in [27], the focus lies on identifying approximate clusters of users and their temporal change. No exact patterns are found, nor is time part of the results obtained with these approaches. [28] uses the MDL principle for monitoring the evolution of a network.

Network motifs as small subgraphs that show the topological properties of the network have also been used by [29] to monitor temporal changes in the structure of an email communication network. They considered z-score of motifs as its significance in the network and trace it over time. The dynamics of a network are studied in [22] by calculating the network motifs frequency over the time.

III. METHODOLOGY

A network shows relations between objects, represented as nodes. Structure and characteristics of a network are technically due to properties of nodes. We address the characterization of network dynamics by tracking the evolution of groups of nodes over time. The constitutive groups of nodes are found in the networks where each group represent a part of characteristics of network. In addition, the dynamics of the network are assessed by regarding the changes happened in the groups.

Our proposed method mainly has two phases, as is illustrated by figure 1. The first one finds the groups of nodes building the networks. In this step basically nodes are clustered regarding their local properties. The second step explores events happening in the networks by deriving a set of rules that explain that evolution.

A. Node label acquisition

Nodes in a network play different roles, holding different positions such as being a hub or a bridge. Labels of nodes are normally pre-defined classes assigned to nodes that represent the role or functionality of nodes in the network. In our method, the label of a node is automatically determined based on its properties in the network. The same label is assigned to the nodes that are in similar position, having similar properties. We assess the distance of two nodes by their properties, rather than using the number of edges between them. Two nodes are close if they have a similar *feature vector*.

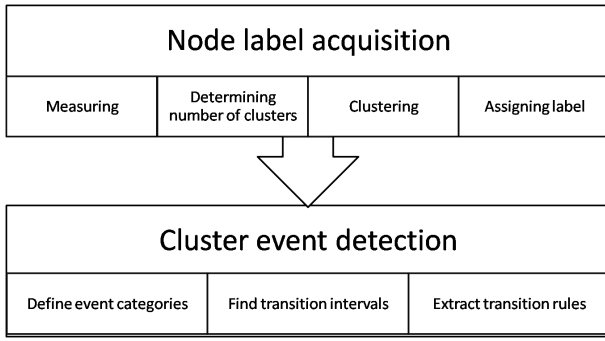


Fig. 1: Proposed methodology

Initially, we need to select a set of local measurements that best characterize nodes in the network structure. Out of a large number of existing properties [9], we employ the same set of metrics used by Costa et al. [11] as the “feature vector” for finding groups of nodes. This feature vector has the advantage of measuring the connectivity of a node in the neighborhood structure. It includes:

- the normalized average degree (r),
- the coefficient variation of the degrees of the immediate neighbors of a node (cv),
- the clustering coefficient (cc),
- the locality index (loc), which is an extension of the matching index and takes into account all the immediate neighbors of each node, instead of individual edges, and
- the normalized node degree (K).

We use multivariate statistics and pattern recognition techniques [30] to find groups of identical nodes. Clustering is a method widely used for finding groups of objects, called clusters, in the dataset such that the objects in the same group are more similar to each other than they are to objects of other groups. We use the well known k-means clustering algorithm [31], which bases its operation on the euclidean distance between nodes. This distance is calculated for every two nodes by considering all five features. Each cluster contains nodes with a similar position in the network regarding their feature vectors. Hence, the same role or label can be assigned to them.

The number of potential groups of nodes in the network is equal to the number of clusters in the dataset. Determining the actual number of groups in a dataset is a fundamental and largely unsolved problem in cluster analysis. We employ the method by [32], since it does not require parametric assumptions, is independent of the method of clustering, and was shown to achieve excellent results. This method uses a theoretic information approach that considers the transformed distortion curve $d_K^{-p/2}$ [32]. “Distortion” is a measure of within cluster dispersion which is a kind of average Mahalanobis distance between the data and the set of cluster centers as a function of the number of clusters, K . This method is called the “jump method”. First, it runs the k-means algorithm for different numbers of clusters, K , and calculates the corresponding distortions, \hat{d}_K . Then it transforms the distortion by power transformation of $y = p/2$, where p is

the number of dimensions in the dataset. The “jumps” in the transformed distortion are calculated by $J_k = \hat{d}_K^{-y} - \hat{d}_{K-1}^{-y}$. Finally, the appropriate number of clusters for the data is equal to $K^* = \argmax_k J_k$.

We use the jump method to determine the number of groups of nodes. We do the clustering on the aggregated dataset that includes feature vectors of every node for the whole lifetime of the network. At the end of this phase, the coherent groups of nodes are derived and labeled. Therefore, a sequence of labels is generated for each node over time, that determines to which cluster a node belongs at each time. In the next phase, we attempt to extract occurring events in the network, considering these sequences.

B. Cluster event detection

Changes in a network are due to some basic events: node addition or deletion, and new edge addition or deletion. These events generate more complex behaviors in networks such as group formation or dissolution. We define five basic types of events for a group according to the changes in its size, that is, the number of constituent nodes. If the size of a group of nodes changes considerably, it shows that properties of a number of nodes has changed and subsequently structure of network is altered. In a certain time interval in the network, a group of nodes can grow, shrink, emerge, dissolve or remain constant. Regarding these behaviors, we defined six event in a network life time as follows:

- **Growth:** A group grows if its size has a constant increasing trend in a time interval.
- **Shrink:** A group shrinks if its size has a constant decreasing trend in a time interval.
- **Emerge:** A new group emerges if its size has a constant increasing trend in a time interval and it does not exist in the previous intervals.
- **Dissolve:** A new group dissolve if its size has a constant decreasing trend in a time interval and it does not exist in the next intervals.
- **Constant:** A new group remains constant if its size does not change. In this case, there might be some nodes leaving or joining although the size of does not change considerably.

There are two primary event categories occurring in the network: shrink and growth. The other ones are specific cases of the original ones. For example, dissolve is a special case of shrink event where the size of the group shrinks to zero. All the occurring events in a network of any of these types are discovered and described by a two step method. In first step, all time intervals where an event occurs in a cluster of nodes are detected, such interval is called transition intervals. Then a set of rules describing the event is generated. Our goal is to discover rules that describe each event, such as rules that can explain the origin of new nodes that have joined a group in a certain time interval.

1) *Transition intervals:* a transition interval is a time interval where a considerable number of nodes leave or join a group. For a given group C_i , the size of group over time

constitutes a time series denoted as $F_i(t), t \in [1, T]$. A transition interval is the subsequence of F_i which holds a constant increasing or decreasing trend. Hence, we extract transition intervals of F_i by segmentation of the time series. Starting from $t = 1$, $F_i(t)$ is approximated by linear regression to find the transition intervals. If the error of the fitted line for a subsequence $F_i[a : b]$ exceeds the threshold, the interval $[a, b]$ breaks to point j where it gives the best approximation for $F_i[a, j - 1], j < b$. The error is measured in terms of sum square of residuals. The threshold is controlled by the maximum number of arbitrary transition intervals. The maximum error is increased until the number of intervals is not more than the defined maximum number of intervals. For this method we can either define the maximum error for the linear regression, or the maximum number of desired intervals. The slope of the fitted line for each segment shows if the interval is increasing or decreasing which respectively determines the growth (emergence) or shrink (dissolve) events.

2) *Transition rules*: having a list of transition intervals for each cluster of nodes, we extract a set of rules to describe how an event happened. A transition rule is in the form of $C_i \rightarrow C_j$ which shows nodes from group i moved to j . The order of transitions is important, since it shows the trend of changes in the network properties. For example, for a sequence of $\{1, 1, 1, 2, 2, 3\}$, the transition rules are $1 \rightarrow 2$ and $2 \rightarrow 3$. We extract these one-step transition rules for a time interval by building a transition matrix in each time interval. The support count of a rule $C_i \rightarrow C_j$ is defined as the number of nodes that go from cluster C_i to C_j in that interval.

In the next section we apply our methodology on three different networks to characterize the dynamics of these networks and evaluate the applicability of our method.

IV. EXPERIMENTS

The proposed methodology for automatically assigning labels to nodes and track their evolution over time was implemented in R. We first describe in some detail the used networks and then we show the obtained results, along with its evaluation.

A. Networks

For our experiments, we used three different real complex networks: a network of the world countries' global trade which we call GDP data [33], a network of USA airports¹ and a co-authorship network obtained from DBLP data [8]. We use the undirected form of these networks. Table I overviews some topological features of the three studied networks.

GDP: The first network is created from the Expanded Trade and GDP Data [33]. The data represents the yearly imports and exports, total trade and gross domestic product (GDP) of 196 countries spanning the 52 years from 1948 till 2000. The time series for each country is the proportion of its share in the global economy according to its GDP for that year. The annual trade network is built by comparing the total trade for

Dataset	Time snapshots	$ V $	$ E $	Node growth rate	Edge growth rate
GDP	53	186	8839	2.47	7.93
USA airports	244	1919	14391	1.64	1.21
DBLP	11	31592	49599	3.4	4.57

TABLE I: Datasets statistics: number of time-snapshots, number of nodes and edges at the final snapshot and node and edge growth rate (ratio between the final and initial time-snapshots).

each country and its trade with each of the other countries. Two countries A and B in the network at time t are connected if the trade between them in year t is more than 10% of either country's total trade in year t .

USA airports: This data is the complete US airport network from 1990 until April of 2011. We constructed monthly networks where two airports are connected if a flight was scheduled between them in that month.

DBLP: This is a co-authorship network from the DBLP data with a yearly time granularity. The nodes are authors that are connected in a certain year if they are co-authors in that year. It includes co-authorship data from 1992 to 2002 [8].

B. Results and evaluation

Here, we present the results of our experiments on the datasets. Table II provides a brief overview of the derived results for the networks, namely the number of clusters, the number of events and the number of rules, extracted in each network.

Dataset	Num. of clusters	Num. of events	Num. of rules
GDP	4	22	66
USA airports	6	52	296
DBLP	8	25	195

TABLE II: Datasets statistics: Number of clusters found, numbers of detected events and number of extracted rules.

This methodology gives us a profile for the network, including information about:

- The set of clusters, existing in the networks;
- The set of events, occurred in the network time span.

Figure 2 shows the profile of the groups found in each network. The profile depicts the values of the feature vector of each group in the network. As explained earlier, the feature vector includes the metrics normalized average degree (r), coefficient variation of the degrees of immediate neighbors (cv), the clustering coefficient (cc), the locality index (loc), and the normalized node degree (K).

For example, in the GDP global trade network of countries, our method found four distinguishable groups of nodes. The first group includes nodes that represent countries with very high degree and many low degree nodes connected to them. Neighbors of these nodes have low degree since the normalized average degree of the immediate neighbors of a node for this group is very low. This means that nodes of group one

¹<http://www.routeviews.org>

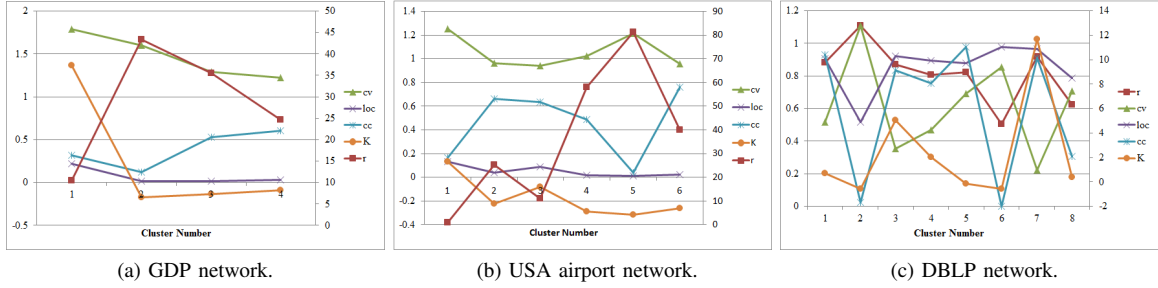


Fig. 2: The feature vector of the clusters in selected networks. The normalized average degree (r) is plotted on the second axes on right side of plot.

behave as hubs in the network, that is, as hub countries in global trade, with commercial transactions with many other countries that have a high variation of degree in neighborhood (cv). According to the value of (loc) and (cc), respectively, the locality index and the clustering coefficient, nodes of this group are highly connected in their neighborhood. United States of America, Canada and France are members of this group.

Dataset	Event	Time interval	Transition rules	Support	Z score
GDP	Emerge (2)	[26,32]	$3 \rightarrow 2$	80	6.3
	Shrink (4)	[23,40]	$4 \rightarrow 3$	100	14.70
	Emerge (3)	[12,22]	$4 \rightarrow 3$	126	8.3
	Shrink (1)	[9,29]	$1 \rightarrow 4$	25	12.1
USA airports	Growth (3)	[142,175]	$1 \rightarrow 3$	1062	3.75
	Growth (6)	[1,79]	$2 \rightarrow 6$	437	7.27
	Shrink (2)	[57,132]	$2 \rightarrow 3$	551	4.79
DBLP	Growth (4)	[1,11]	$0 \rightarrow 4$	6710	1.2
	Growth (7)	[6,11]	$0 \rightarrow 7$	196	0.2
	Shrink (3)	[1,7]	$3 \rightarrow 0$	726	1.3

TABLE III: Description of some extracted events in the networks. Numbers in the parenthesis denotes the cluster number, holding the events

After this first clustering step, we continued with the second phase. Table III shows some of the extracted events. Besides the frequency count of every event in each dataset, and in order to assert their significance, we compared the result with randomized sequences. We built these random sequences by shuffling the order of cluster membership of each node in the network. This way, the random dataset has the same number of nodes and cluster types. Since we have a dataset of cluster sequences for each network, we built 10 datasets of random sequences with the size as the number of nodes. We calculated average and standard deviation of frequencies for each rule of events in the 10 datasets. Finally, we calculated the Z-score for the significance of each rule as compared to the randomized form. We show in the table the rules with the highest z-score.

Figure 3 graphically illustrates some of the events happening in each network in terms of transition of clusters. Each graph includes all transitions of clusters in the specified time interval. The label of nodes is the cluster number and the color of nodes shows the type of event happened to the cluster.

For example, the first event (a) in the GDP network describes the emergence of cluster 3 in the [12,22] time interval.

This graph says that nodes that constitute cluster 3 either come from cluster 4 or are the new nodes (cluster 0), just joining the network. By observing Table III, we can see that the main reason for this event is the transition from cluster 4. The same type of interpretation could be applied for the two other networks.

V. CONCLUSIONS

Many networks are intrinsically dynamic and evolve over time. Discovering topological features in these networks is far from an easy task. In this work, we proposed a network characterization method that considers both a static and a dynamic point of view. It is a two phase methodology that automatically assigns labels to nodes of the network based on their local properties and extracts events happening during the evolution of network. The static view provides a general description of the network through label assignment to groups of nodes. Each group in the network is well characterized by the corresponding feature vector profiling. From a dynamic point of view, the methodology discovers five categories of events for each cluster, emerge, growth, constant, shrink and dissolve. The extracted events are described by some rules that depict the reason of each event and the flow of transition between clusters.

We applied our method to three real data networks to demonstrate and assess its capabilities. It successfully clusters nodes in groups performing a similar role in the network, labels the groups and, through the second phase of the methodology, derives events and their explanation in the form of transition rules. The rules show node transitions between clusters in the time interval that event happened. We evaluated the extracted rules in term of z-score.

Future research will be pursued to extend this methodology, so that we do not just look at individual nodes but subgraphs in the network. In particular, given our prior work on efficient methods for motifs discovery [34], [35], we are specially concerned with using subgraph motifs as a metric for network characterization, and then studying network evolution based on such larger entities.

ACKNOWLEDGMENTS

This work is in part funded by the ERDF/COMPETE Programme and by FCT within project FCOMP-01-0124-FEDER-022701. S. Choobdar and P. Ribeiro are funded by FCT grants

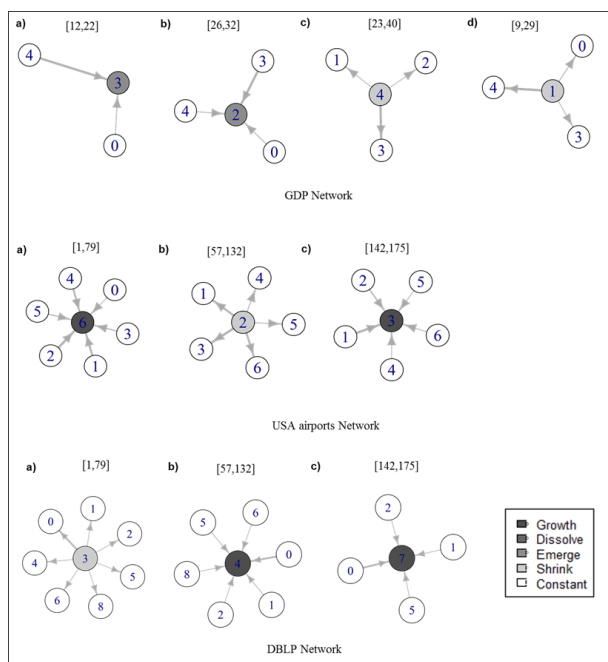


Fig. 3: Some of the events happening in the networks.

SFRH/BD/72697/2010 and SFRH/BPD/81695/2011. The authors are also thankful to Maria Eduarda Silva for the fruitful discussions and suggestions.

REFERENCES

- [1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *PNAS*, vol. 99, no. 12, pp. 7821–7826, June 2002.
- [2] S. E. Schaeffer, "Graph clustering," *Computer Science Review*, vol. 1, no. 1, pp. 27–64, 2007.
- [3] S. Macskassy and F. Provost, "Classification in networked data: A toolkit and a univariate case study," *The Journal of Machine Learning Research*, vol. 8, pp. 935–983, 2007.
- [4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, vol. 298, no. 5594, pp. 824–827, October 2002.
- [5] X. Yan and J. Han, "gspan: Graph-based substructure pattern mining," in *Proceedings of the 2002 IEEE International Conference on Data Mining*, ser. ICDM '02, 2002, pp. 721–.
- [6] J. Huan, W. Wang, and J. Prins, "Efficient mining of frequent subgraphs in the presence of isomorphism," in *Proceedings of the Third IEEE International Conference on Data Mining*, ser. ICDM '03, 2003, pp. 549–.
- [7] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 177–187.
- [8] M. Berlingerio, F. Bonchi, B. Bringmann, and A. Gionis, "Mining graph evolution rules," in *Machine Learning and Knowledge Discovery in Databases*. Springer, September 2009, pp. 115–130.
- [9] L. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, "Characterization of complex networks: A survey of measurements," *Advances In Physics*, vol. 56, p. 167, 2007.
- [10] S. Choobdar, F. Silva, and P. Ribeiro, "Network node label acquisition and tracking," in *Progress in Artificial Intelligence, 15th Portuguese Conference on Artificial Intelligence - EPIA'11, Lisbon, Springer, LNAI 7026*, October 2011, pp. 418–430.
- [11] L. Costa, F. Rodrigues, C. Hilgetag, and M. Kaiser, "Beyond the average: detecting global singular nodes from local features in complex networks," *EPL (Europhysics Letters)*, vol. 87, p. 18008, 2009.
- [12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.
- [13] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.
- [14] X. Xu, N. Yuruk, Z. Feng, and T. Schweiger, "Scan: a structural clustering algorithm for networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 824–833.
- [15] Y. Zhou, H. Cheng, and J. Yu, "Graph clustering based on structural/attribute similarities," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 718–729, 2009.
- [16] Y. Tian, R. Hankins, and J. Patel, "Efficient aggregation for graph summarization," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 567–580.
- [17] B. Taskar, P. Abbeel, and D. Koller, "Discriminative probabilistic models for relational data," in *18th Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 895–902.
- [18] D. Jensen, J. Neville, and B. Gallagher, "Why collective inference improves relational classification," in *Proceedings of the 10th KDD*, 2004.
- [19] B. Gallagher and T. Eliassi-Rad, "Leveraging label-independent features for classification in sparsely labeled networks: an empirical study," in *Proceedings of the 2nd international conference on Advances in social network mining and analysis*, ser. SNAKDD'08. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 1–19.
- [20] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 44–54.
- [21] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 462–470.
- [22] D. Braha and Y. Bar-Yam, "Time-dependent complex networks: Dynamic centrality, dynamic motifs, and cycles of social interactions," in *Adaptive Networks*, 2009, vol. 51, pp. 39–50.
- [23] C. Tantipathanandh, T. Berger-Wolf, and D. Kempe, "A framework for community identification in dynamic social networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 717–726.
- [24] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "Graphscope: parameter-free mining of large time-evolving graphs," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 687–696.
- [25] C. Aggarwal and P. Yu, "Online analysis of community evolution in data streams," in *Proceedings of the 15th SIAM International Conference on Data Mining*. Society for Industrial Mathematics, 2005, p. 56.
- [26] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: dynamic tensor analysis," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 374–383.
- [27] N. Vanetik, S. Shimony, and E. Gudes, "Support measures for graph data," *Data Mining and Knowledge Discovery*, vol. 13, no. 2, pp. 243–260, 2006.
- [28] J. Ferlez, C. Faloutsos, J. Leskovec, D. Mladenec, and M. Grobelsnik, "Monitoring network evolution using MDL," in *IEEE 24th International Conference on Data Engineering*, 2008, pp. 1328–1330.
- [29] K. Juszczyszyn, P. Kazienko, K. Musial, and B. Gabrys, "Temporal Changes in Connection Patterns of an Email-Based Social Network," in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 3, 2008, pp. 9–12.
- [30] R. Johnson and D. Wichern, *Applied multivariate statistical analysis*. Prentice Hall Upper Saddle River, NJ, 2002, vol. 5.
- [31] J. Hartigan and M. Wong, "A k-means clustering algorithm," *Journal of the Royal Statistical Society C*, vol. 28, no. 1, pp. 100–108, 1979.
- [32] C. Sugar and G. James, "Finding the number of clusters in a dataset," *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 750–763, 2003.
- [33] K. Gleditsch, "Expanded trade and GDP data," *Journal of Conflict Resolution*, vol. 46, no. 5, p. 712, 2002.
- [34] P. Ribeiro and F. Silva, "G-tries: an efficient data structure for discovering network motifs," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 1559–1566.
- [35] P. Ribeiro, F. Silva, and M. Kaiser, "Strategies for network motifs discovery," in *Proceedings of the 2009 Fifth IEEE International Conference on e-Science*, ser. E-SCIENCE '09, 2009, pp. 80–87.